

## **Análise Automática de Feedback em Ambientes de Aprendizagem Online**

**Anderson Pinheiro Cavalcanti<sup>1</sup>, Rafael Ferreira Mello<sup>2</sup>, Péricles Miranda<sup>2</sup>, Fred Freitas<sup>1</sup>**

<sup>1</sup>Centro de Informática – Universidade Federal de Pernambuco (UFPE)

<sup>2</sup>Departamento de Computação – Universidade Federal Rural de Pernambuco (UFRPE)

{apc, fred}@cin.ufpe.br, {rafael.mello, pericles.miranda}@ufrpe.br

**Abstract.** *Feedback is a very important component in the teaching-learning process. In distance courses, feedback is one of the most used resources in the interaction between teacher and student. However, due to the significant growth in the number of students taking distance courses, it is difficult for instructors to provide high quality feedback. In this context, this article aims to propose an approach to automatically analyze the feedback provided by teachers in online courses. For that, linguistic resources and machine learning algorithms were used. The results found show a great improvement in accuracy and Kappa when compared to a previous study.*

**Resumo.** *O feedback é um componente muito importante no processo de ensino-aprendizagem. Em cursos a distância o feedback é um dos recursos mais utilizados na interação entre professor e aluno. No entanto, devido ao crescimento significativo da quantidade de alunos em cursos a distância, é difícil para os instrutores fornecerem um feedback de alta qualidade. Nesse contexto, este artigo tem como objetivo propor uma abordagem para analisar automaticamente os feedbacks fornecidos por professores em cursos online. Para isso, foram utilizados recursos linguísticos e algoritmos de aprendizagem de máquina. Os resultados encontrados mostram uma grande melhoria na acurácia e Kappa quando comparado com um estudo anterior.*

### **1. Introdução**

O feedback é um fator essencial no processo de aprendizagem, pois permite aos alunos identificar lacunas no aprendizado, pode ajudar os alunos a identificar áreas de melhoria em seus conhecimentos ou habilidades, e refletir em suas estratégias de aprendizagem [Sadler 1989]. Além disso, os professores usam o feedback para identificar as necessidades dos alunos, para que possam adaptar seus métodos e conteúdo com base nessas necessidades [Langer 2011]. A literatura demonstra os benefícios e o impacto positivo que um bom feedback causa no aprendizado do aluno [Nicol and Macfarlane-Dick 2006], mas também ressaltam que um feedback ruim pode desmotivar e até levar a evasão do aluno [Hattie and Timperley 2007].

Outro fator relevante neste contexto é que o uso de Ambientes Virtuais de Aprendizagem (AVAs) tem aumentado nos últimos anos devido ao uso das tecnologias de informação e comunicação como ferramenta de apoio educacional

[Correia and dos Santos 2013]. Esses ambientes possuem ferramentas que permitem uma grande interação entre professores e alunos, como por exemplo: chat, fórum, wiki, entre outras. Entre elas, há a ferramenta de envio de atividades em que os alunos podem enviar suas respostas às atividades em andamento. Em geral, esse recurso de avaliação é o principal espaço onde os instrutores enviam feedback [Coates et al. 2005]. No entanto, devido ao crescimento significativo da quantidade de alunos nos AVAs que precisam de um envolvimento contínuo, é difícil para os instrutores fornecerem um feedback personalizado e de alta qualidade [Espasa and Meneses 2010].

Com o objetivo de reduzir o esforço do instrutor e melhorar a qualidade do feedback, pesquisas recentes trazem algumas abordagens para automatizar o feedback para os alunos [Marin et al. 2017, Belcadhi 2016]. Entretanto, esses trabalhos abordam o feedback em um contexto específico, como por exemplo cursos de programação introdutórios.

Nesse contexto, este artigo propõe uma abordagem para analisar automaticamente os feedbacks enviados pelos instrutores com base em alguns princípios de boas práticas de feedback [Nicol and Macfarlane-Dick 2006]. Para isso, foram analisados os algoritmos XGBoost e AdaBoost para extrair indicadores que podem levar à melhoria do feedback enviado aos alunos. A abordagem proposta alcançou resultados até quatro vezes melhores quando comparado aos resultados de trabalhos anteriores.

## 2. Trabalhos Relacionados

O feedback é um componente essencial da aprendizagem, uma vez que direciona os alunos para o tipo apropriado de estudo ou prática. Esse tópico é uma parte fundamental do processo de aprendizado do aluno e vem sendo abordado em diversas pesquisas nas últimas duas décadas [Nicol and Macfarlane-Dick 2006, Jeria and Villalon 2017]. Embora muitas pesquisas relatem efeitos positivos do feedback, nem todo feedback é igualmente eficaz. Por exemplo, no estudo de [Burke 2009] foram levantados alguns fatores de insatisfação do aluno com relação ao feedback recebido. Entre esses fatores, estão o comprimento (o feedback é muito breve), a polaridade (o feedback é muito negativo) e a complexidade (é muito difícil decifrar ou entender o feedback).

A literatura também oferece recomendações de boas práticas de feedback. [Nicol and Macfarlane-Dick 2006] propôs um modelo conceitual de autorregulação com base em uma revisão da literatura de pesquisa sobre avaliação formativa e feedback. A ideia principal do trabalho é identificar como os processos formativos de avaliação e feedback podem ajudar a promover a autorregulação. Com base no modelo conceitual, foram definidos sete princípios de boas práticas de feedback que o professor pode usar para refletir sobre o projeto e avaliar seus próprios procedimentos de feedback. Abaixo estão as sete boas práticas<sup>1</sup> de feedback [Nicol and Macfarlane-Dick 2006]. De acordo com os autores, boas práticas de feedback são amplamente definidas como qualquer coisa que possa fortalecer a capacidade dos alunos de autorregular seu desempenho.

- **BP 1:** Ajuda a esclarecer o que é um bom desempenho (metas, critérios, padrões esperados);
- **BP 2:** Facilita o desenvolvimento da autoavaliação (reflexão) na aprendizagem;
- **BP 3:** Fornece informações de alta qualidade aos alunos sobre seu aprendizado;

---

<sup>1</sup>Foi utilizado o acrônimo BP para Boas Práticas

- **BP 4:** Incentiva o diálogo entre professores e colegas sobre o aprendizado;
- **BP 5:** Incentiva crenças motivacionais positivas e autoestima;
- **BP 6:** Oferece oportunidades para fechar a lacuna entre o desempenho atual e o desejado;
- **BP 7:** Fornece informações aos professores que podem ser usadas para ajudar a moldar o ensino;

Na literatura também é possível encontrar alguns trabalhos que analisam automaticamente o feedback. [Maitra et al. 2018] propôs um classificador automático para investigar o feedback usando o classificador Naïve Bayes. O autor classificou o feedback fornecido para cada aluno em duas classes, válidas ou inválidas, com base em uma medida de feedback proposta. O classificador leva em consideração a contribuição independente de diferentes recursos do aluno (ou seja, esforço, formação acadêmica, resultados do curso alcançados), no feedback fornecido pelo instrutor. A abordagem foi avaliada usando 1.000 extratos de feedback de uma instituição de ensino superior indiana. A principal desvantagem dessa abordagem foi a falta de detalhes sobre a influência dos recursos utilizados e as implicações adicionais de sua abordagem.

Os métodos de mineração de texto também foram usados para fornecer feedback automático aos alunos. [Akçapınar 2015] visava reduzir o plágio dos alunos nas tarefas *online*, fornecendo feedback automatizado com base na análise de mineração de texto. Antes do feedback automático, 81,4% dos estudantes cometeram plágio. Após o feedback, 83% dos estudantes reduziram as tentativas pós-plágio. O número de não plagiadores aumentou 42,37% após o fornecimento de feedback.

Por fim, o estudo de [Cavalcanti et al. 2019] teve como objetivo analisar a qualidade do feedback extraído das avaliações coletadas em um curso *online* oferecido em uma instituição de ensino superior brasileira. Um classificador *Random Forest* foi usado para classificar as mensagens de feedback, verificando se os textos seguiam as boas práticas propostas por [Nicol and Macfarlane-Dick 2006]. Os autores avaliaram o sistema usando um conjunto de dados que continha 1.000 exemplos de mensagens de feedback. O classificador alcançou 0,75 e 0,20 para as medidas de acurácia e kappa, respectivamente.

Dessa forma, este trabalho se difere dos trabalhos mencionados acima em duas principais contribuições: (1) Utilizar algoritmos de aprendizagem de máquina mais recentes para classificar os feedback; e (2) Analisar as características mais importantes para cada classificador.

### 3. Perguntas de Pesquisa

Esse trabalho tem como objetivo automatizar a classificação de feedback de cursos a distância seguindo as boas práticas de feedback propostas por [Nicol and Macfarlane-Dick 2006]. Dessa forma, nossa primeira questão de pesquisa é:

**PERGUNTA DE PESQUISA 1:** *Qual classificador obtém o melhor desempenho na classificação de feedback de acordo com os critérios de boas práticas?*

Além de verificar qual classificador obtém melhores resultados, também propomos a análise de quais as características são mais relevantes para esses classificadores.

**PERGUNTA DE PESQUISA 2:** *Quais as características mais importantes na classificação dos feedback?*

## 4. Método

### 4.1. Dados

O conjunto de dados utilizado neste trabalho é o mesmo utilizado por [Cavalcanti et al. 2019, Cavalcanti et al. 2020] e foi obtido de um AVA usado em cursos *online* oferecidos por uma universidade pública brasileira. O conjunto de dados é composto por mensagens de texto que o instrutor enviou como feedback para as atividades de seus alunos enviadas pelo AVA. Foram coletados um total de 1.000 feedbacks dos cursos de biologia e literatura divididos em 41 mensagens para biologia e 959 para literatura. Todo o conjunto de dados foi anotado por especialistas que analisaram se o feedback seguia as boas práticas propostas por [Nicol and Macfarlane-Dick 2006]. O conjunto de dados está dividido em duas classes: classe 0: se o feedback não pertencer a nenhuma das boas práticas; classe 1: se o feedback pertencer a pelo menos uma das sete boas práticas. A Tabela 1 mostra a quantidade de feedback por classe.

**Tabela 1. Distribuição do banco de dados por classe.**

Classe	Descrição	Quantidade
0	Não possui boa prática	222
1	Possui 1 ou mais boas práticas	778
<b>Total</b>		<b>1000</b>

### 4.2. Extração de Características

A extração de características é o processo que transforma o texto do feedback em dados numéricos. Nesse trabalho foram utilizados dois sistemas que extraem características linguísticas do texto em conjunto com 4 características adicionais utilizando mineração de texto. Todos esses recursos são detalhados a seguir.

O LIWC (*Linguistic Inquiry and Word Count*) extrai um grande número de contagens de palavras que são indicativas de diferentes processos psicológicos, como afetivos, cognitivos, sociais e perceptivos [Tausczik and Pennebaker 2010]. O LIWC possui 127.149 entradas, nas quais cada entrada pode ser atribuída a uma ou mais categorias. O núcleo desta ferramenta é um recurso lexical, mais conhecido como dicionário LIWC, que também foi disponibilizado no idioma português [Balage Filho et al. 2013]. Um total de 64 recursos LIWC foram extraídos no estudo.

Coh-Metrix é uma ferramenta computacional que analisa muitas palavras, frases e textos com várias frases diferentes. Coh-Metrix fornece várias medidas de coerência do texto, complexidade linguística, legibilidade do texto e uso da categoria lexical [McNamara et al. 2014]. A versão em português do Coh-Metrix possui 48 medidas diferentes.

Além dos recursos LIWC e Coh-Metrix, foram extraídos do texto: o número de entidades nomeadas, polaridade de sentimento (positiva, negativa e neutra), presença de saudação (“Bom dia”, “Boa tarde”, “Como vai”) e expressões positivas (“Muito bem”, “Excelente”, “Bom trabalho”). Esses recursos também foram propostos em trabalhos de classificação de texto educacional [Ferreira-Mello et al. 2019]. Para isso, foi utilizado o SentiLex-PT, que é um léxico criado explicitamente para a análise de sentimentos

e opiniões sobre entidades humanas em textos da língua portuguesa. Cada entrada do SentiLex-PT contém a polaridade de uma palavra, que pode ser positiva, neutra ou negativa [Carvalho and Silva 2015].

Por fim, a biblioteca spaCy<sup>2</sup> foi usada para extrair o número de entidades nomeadas. O número final de características foram 116, com 64 do LIWC, 48 do Coh-Metrix e 4 características adicionais.

### 4.3. Processamento de Dados e Classificação

No aprendizado de máquina, para evitar superestimar o desempenho do modelo, é necessário dividir o conjunto de dados em conjuntos de treinamento e teste [Friedman et al. 2001]. O conjunto de dados foi dividido em conjuntos de treinamento (80%) e teste (20%). Além disso, os dados estão desbalanceados, como mostra a Tabela 1, a classe 1 tem quase 4 vezes mais exemplos que a classe 0. O desbalanceamento de classe pode ter efeitos muito negativos nos resultados das análises de classificação. Assim, o algoritmo SMOTE foi usado para resolver o desequilíbrio do conjunto de dados. Esse algoritmo cria pontos de dados sintéticos adicionais como uma combinação linear dos pontos de dados existentes [Chawla et al. 2002].

Para classificar os feedback foram utilizados 2 algoritmos de aprendizagem de máquina baseados em árvores de decisão, o AdaBoost<sup>3</sup> e o XGBoost<sup>4</sup>. Esses classificadores possuem alguns parâmetros de treinamento que devem ser definidos. Neste trabalho foi analisado o parâmetro *n<sub>tree</sub>* que significa o número de árvores geradas pelo algoritmo. Para obter o parâmetro ideal, foi utilizada a técnica validação cruzada de 10 vezes nos dados de treinamento e depois aplicado o modelo com melhor acurácia nos dados de teste. Os 2 classificadores também fornecem métodos para analisar a importância das características na classificação, que foram utilizados para abordar a questão de pesquisa 2 no presente estudo. Uma medida popular para calcular a importância do recurso é o *Mean Decrease Gini (MDG)*, que se baseia na redução na medida de impureza de Gini [Breiman 2001]. Neste artigo, foi utilizado o MDG para avaliar a relevância de diferentes características para o resultado dos classificadores.

#### 4.3.1. AdaBoost

Adaboost (*Adaptive Boosting*) é um algoritmo de classificação que utiliza a técnica *Boosting* para classificar os exemplos. *Boosting* é uma técnica que combina os classificadores gerados por um mesmo algoritmo de aprendizado e com isso consegue formar um classificador “forte” com base em classificadores mais simples, ditos “fracos”. O AdaBoost mantém um conjunto de pesos sobre os exemplos de treinamento. Em cada iteração, o algoritmo de aprendizado é chamado para minimizar o erro ponderado no conjunto de treinamento. Esse erro é utilizado para ponderar os pesos nos exemplos de treinamento. Dessa forma, o algoritmo coloca mais peso nos exemplos classificados incorretamente e menos peso nos exemplos classificados corretamente [Dietterich 2000].

---

<sup>2</sup><https://spacy.io>

<sup>3</sup><https://scikit-learn.org/>

<sup>4</sup><https://xgboost.readthedocs.io/>

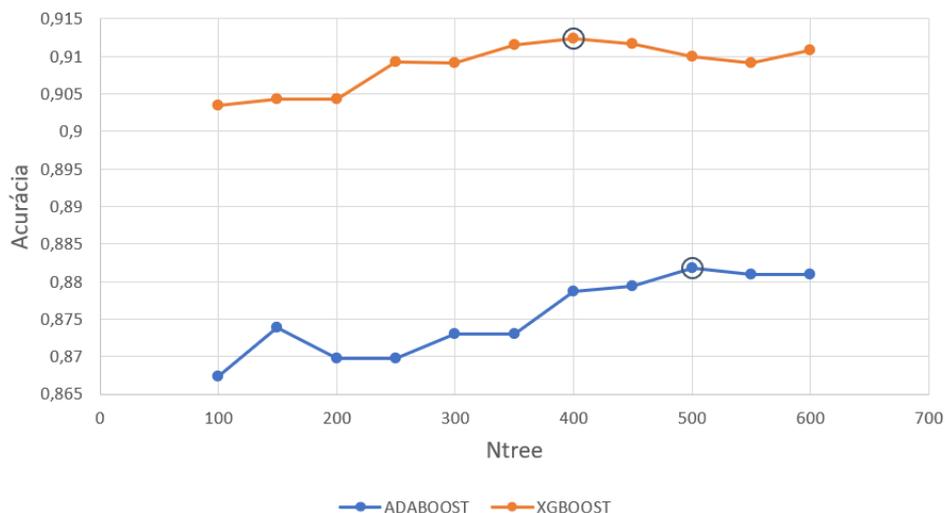
### 4.3.2. XGBoost

O algoritmo XGBoost é um modelo de aprendizado de máquina escalonável para o aprimoramento de árvore baseado em árvores de Decisão de Intensidade de Gradiente. O XGBoost fornece um aumento de árvore paralelo (também conhecido como GBDT - *Gradient Boosted Decision Tree*) que resolve muitos problemas de ciência de dados de maneira rápida e precisa. Diferente dos métodos tradicionais de reforço que pesam amostras positivas e negativas, o GBDT faz convergência global do algoritmo seguindo a direção do gradiente negativo [Chen and Guestrin 2016].

## 5. Resultados

### 5.1. Análise dos classificadores

Inicialmente, os classificadores foram aplicados no conjunto de treinamento utilizando a técnica de validação cruzada 10 vezes para diferentes valores de *ntree*. A Figura 1 mostra os resultados obtidos para os dois classificadores.



**Figura 1. Resultados dos classificadores no conjunto de treinamento.**

A melhor acurácia do AdaBoost foi de 0,88 (Kappa = 0,76) obtida com *ntree* = 500. Para o XGBoost a melhor acurácia foi de 0,91 (Kappa = 0,82) com *ntree* = 400. Pela Figura 1 é possível perceber que o XGBoost obteve a maior acurácia em todos os casos quando comparado com o AdaBoost. A diferença entre a pior e a melhor acurácia foi de 0,0145 e 0,0089 para AdaBoost e XGBoost, respectivamente. Isso mostra que a otimização do parâmetro *ntree* é essencial no desempenho final do classificador.

O melhor modelo obtido no conjunto de treinamento foi aplicado no conjunto de teste. O AdaBoost obteve 0,89 de acurácia (Kappa = 0,78) e o XGBoost obteve 0,91 de acurácia (Kappa = 0,82). A Tabela 2 mostra os resultados obtidos e compara com o resultado obtido por [Cavalcanti et al. 2019].

Da mesma forma que no conjunto de treinamento, o XGBoost obteve o melhor resultado comparado ao AdaBoost e também obteve um aumento substancial na acurácia e principalmente no Kappa quando comparado ao resultado obtido por [Cavalcanti et al. 2019].

**Tabela 2. Comparação resultados na base de teste.**

	Acurácia	Kappa
AdaBoost	0,89	0,78
XGBoost	<b>0,91</b>	<b>0,82</b>
[Cavalcanti et al. 2019]	0,75	0,2

## 5.2. Análise top-10 características

Assim como trabalhos recentes, [Sirotheau et al. 2019], esse estudo também analisou as contribuições das diferentes características no desempenho final do classificador. Como mencionado anteriormente, foi utilizada a medida *Mean Decrease Gini (MDG)* para definir o grau de relevância de uma característica [Breiman 2001]. As Tabelas 3 e 4 mostram as 10 características mais importantes para os classificadores AdaBoost e XGBoost, respectivamente.

**Tabela 3. Top-10 de características mais importantes do AdaBoost.**

#	Característica	Descrição	MDG
1	cw_frequencies:cw_freq,	Frequência de palavras de conteúdo	4,20
2	liwc.we	Contagem de pronomes na 1 <sup>a</sup> pessoa do plural	3,60
3	liwc.posemo	Contagem de palavras positivas	3,60
4	liwc.number	Contagem de números	2,80
5	conn:tmp_neg_conn_inc	Incidência de conectivos de negação	2,60
6	liwc.certain	Contagem de palavras de certeza	2,60
7	liwc.friend	Contagem de palavras de amizade	2,60
8	liwc.excl	Contagem de pontos de exclamação	2,40
9	liwc.affect	Contagem de palavras de afeto	2,40
10	liwc.auxverb	Contagem de verbos auxiliares	2,20

A característica mais importante do AdaBoost foi a frequência de palavras de conteúdo (substantivos, verbos, adjetivos e advérbios). Essa característica pode estar diretamente relacionada ao tamanho do texto fornecendo exemplos para o aluno (usando substantivos) ou elogiando a atividade do aluno (usando adjetivos) seguindo as boas práticas 3 e 5 [Nicol and Macfarlane-Dick 2006], respectivamente. Através da Tabela 3 podemos verificar também que as características *liwc.posemo*, *liwc.friend* e *liwc.affect* podem estar relacionadas a boa prática 5, cujo objetivo é fornecer crenças motivacionais e positivas para o aluno.

A característica mais importante do XGBoost foi a proporção de sentenças adjacentes que compartilham palavras de conteúdo (substantivos, verbos, adjetivos e advérbios). Da mesma forma que o AdaBoost, essa característica em conjunto com *basic\_counts:adverbs*, *liwc.auxverb*, *bc:words\_per\_sentence*, *tokens.ttr* e *cw\_frequencies:cw\_freq* estão relacionadas ao tamanho do texto fornecido pelo professor. Isso significa que o professor forneceu informações de alta qualidade aos alunos

**Tabela 4. Top-10 de características mais importantes do XGBoost.**

#	Característica	Descrição	MDG
1	coreference:adj_cw_ovl	Palavras de conteúdo sobrepostas em frases adjacentes	7,86
2	basic_counts:adverbs	Contagem básica de advérbios	5,43
3	liwc.auxverb	Contagem de verbos auxiliares	4,46
4	constituents:np_incidence	Incidência de Sintagmas	4,23
5	bc:words_per_sentence	Contagem básica de palavras por sentença	4,15
6	anaphoras:anaphoric_refs	Referência anafóricas	3,02
7	cw_frequencies:cw_freq	Frequência de palavras de conteúdo	2,76
8	liwc.we	Contagem de pronomes na 1ª pessoa do plural	2,27
9	liwc.conj	Contagem de conjunções	1,97
10	tokens:ttr	Número de palavras divididas pelo número de tokens	1,92

usando exemplos ou dicas para melhorar as próximas atividades e dessa forma seguiu as boas práticas 3, 5 e 6 [Nicol and Macfarlane-Dick 2006].

## 6. Discussão

Este trabalho analisou automaticamente se os textos de feedback fornecidos por professores em cursos a distância seguiam as boas práticas de feedback propostas por [Nicol and Macfarlane-Dick 2006]. A 1ª pergunta de pesquisa teve como objetivo analisar qual classificador obtém a melhor acurácia para classificar os feedbacks. Foram analisados os classificadores AdaBoost e XGBoost (ambos utilizam árvores de decisão) e comparados com um estudo anterior que utilizou o Random Forest. O classificador XGBoost obteve o melhor resultado em ambos os conjuntos de treinamento e teste comparado ao AdaBoost. Os resultados também mostraram que tanto o AdaBoost quanto o XGBoost conseguiram obter uma melhor acurácia e Kappa comparado ao trabalho de [Cavalcanti et al. 2019] com uma melhoria de 18,66% e 21,33% com relação a acurácia e 290% e 310% com relação ao Kappa, respectivamente.

Além de analisar a acurácia dos classificadores, a questão de pesquisa 2 tinha como objetivo analisar as características mais relevantes no processo de classificação dos feedbacks. Como os classificadores são baseados em árvores de decisão, é possível obter as características que são mais relevantes para cada classificador. Através dos resultados obtidos foi possível verificar relações entre as características mais relevantes e as boas práticas propostas por [Nicol and Macfarlane-Dick 2006]. A abordagem proposta, baseada principalmente nos sistemas Coh-Metrix e LIWC, mostrou que é possível fornecer um sistema totalmente automatizado para auxiliar na verificação se o texto do professor segue as boas práticas.

## 7. Limitações e trabalhos futuros

Uma das limitações desse estudo é relacionado ao conjunto de dados. Os textos obtidos são de apenas dois cursos a distância diferentes (Literatura e Biologia) que foram forne-

cidos através da ferramenta de envio de atividades. Dessa forma, os textos podem não ser totalmente representativos com relação ao tipo de feedback fornecido em cursos a distância. Além disso, o tamanho do conjunto de dados e o desbalanceamento das classes podem afetar o desempenho do classificador. Como trabalhos futuro pretendemos aumentar o conjunto de dados com textos de outros cursos e também criar uma ferramenta online que analise textos de feedback.

## Referências

- [Akçapınar 2015] Akçapınar, G. (2015). How automated feedback through text mining changes plagiaristic behavior in online assignments. *Computers & Education*, 87:123–130.
- [Balage Filho et al. 2013] Balage Filho, P. P., Pardo, T. A. S., and Aluísio, S. M. (2013). An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pages 215–219. Sociedade Brasileira de Computação.
- [Belcadhi 2016] Belcadhi, L. C. (2016). Personalized feedback for self assessment in lifelong learning environments based on semantic web. *Computers in Human Behavior*, 55:562–570.
- [Breiman 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- [Burke 2009] Burke, D. (2009). Strategies for using feedback students bring to higher education. *Assessment & Evaluation in Higher Education*, 34(1):41–50.
- [Carvalho and Silva 2015] Carvalho, P. and Silva, M. J. (2015). Sentilex-pt: Principais características e potencialidades. *Oslo Studies in Language*, 7(1).
- [Cavalcanti et al. 2020] Cavalcanti, A. P., Diego, A., Mello, R. F., Mangaroska, K., Nascimento, A., Freitas, F., and Gašević, D. (2020). How good is my feedback? a content analysis of written feedback. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 428–437.
- [Cavalcanti et al. 2019] Cavalcanti, A. P., Mello, R. F., Rolim, V., André, M., Freitas, F., and Gašević, D. (2019). An analysis of the use of good feedback practices in online learning courses. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, volume 2161, pages 153–157. IEEE.
- [Chawla et al. 2002] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- [Chen and Guestrin 2016] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.
- [Coates et al. 2005] Coates, H., James, R., and Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary education and management*, 11:19–36.
- [Correia and dos Santos 2013] Correia, R. L. and dos Santos, J. G. (2013). A importância da tecnologia da informação e comunicação (tic) na educação a distância (ead) do ensino superior (ies). *Revista Aprendizagem em EAD*, 2(1).

- [Dietterich 2000] Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer.
- [Espasa and Meneses 2010] Espasa, A. and Meneses, J. (2010). Analysing feedback processes in an online teaching and learning environment: an exploratory study. *Higher education*, 59(3):277–292.
- [Ferreira-Mello et al. 2019] Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., and Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(6):e1332.
- [Friedman et al. 2001] Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA.
- [Hattie and Timperley 2007] Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1):81–112.
- [Jeria and Villalon 2017] Jeria, H. and Villalon, J. (2017). Incorporating open education resources into computer supported marking tool to enhance formative feedback creation. In *Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on*, pages 256–260. IEEE.
- [Langer 2011] Langer, P. (2011). The use of feedback in education: a complex instructional strategy. *Psychological reports*, 109(3):775–784.
- [Maitra et al. 2018] Maitra, S., Madan, S., Kandwal, R., and Mahajan, P. (2018). Mining authentic student feedback for faculty using naïve bayes classifier. *Procedia computer science*, 132:1171–1183.
- [Marin et al. 2017] Marin, V. J., Pereira, T., Sridharan, S., and Rivero, C. R. (2017). Automated personalized feedback in introductory java programming moocs. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 1259–1270. IEEE.
- [McNamara et al. 2014] McNamara, D. S., Graesser, A. C., McCarthy, P. M., and Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- [Nicol and Macfarlane-Dick 2006] Nicol, D. J. and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2):199–218.
- [Sadler 1989] Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2):119–144.
- [Sirotheau et al. 2019] Sirotheau, S., Santos, J., Favero, E., and de Freitas, S. N. (2019). Avaliação automática de respostas discursivas curtas baseado em três dimensões linguísticas. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 1551.
- [Tausczik and Pennebaker 2010] Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.